PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-250938
(43)Date of publication of application : 14.09.2000

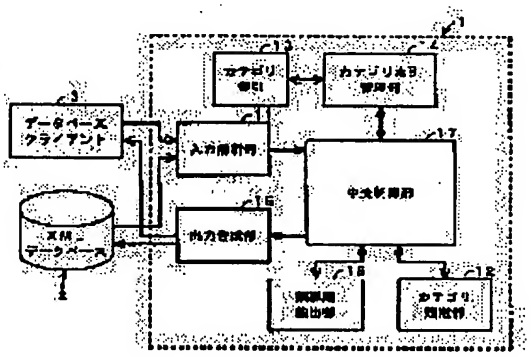
(51)Int.Cl. G06F 17/30

(21)Application number : 11-054960
(22)Date of filing : 03.03.1999

(71)Applicant : KDD CORP
(72)Inventor : ONO TOSHIHIRO
NISHIYAMA SATOSHI
OBANA SADAO

(54) RETRIEVAL DEVICE FOR XML DOCUMENT

(57)Abstract:
PROBLEM TO BE SOLVED: To enable a user to effectively retrieve an XML(extensible markup language) data base having the DTD(document-types tag definitions) which are structurally different from each other and semantically similar to each other with no consciousness of the difference of DTD.
SOLUTION: An input analysis part 11 extracts an element name of an input retrieval expression which is produced by a data base client 3. A central control part 17 acquires a near-synonym of the extracted element name via a near- synonym extraction part 15 and compares this near-synonym with an element name that is stored in a category analogizing part 12 to select the element names which are coincident with each other. Then the part 17 produces an output retrieval expression by means of the selected element name and retrieves an XML data base 2 by using the output retrieval expression. The retrieval result of the data base 2 is notified to the client 3 via the parts 11 and 17 and an output synthesizing part 16.



LEGAL STATUS

[Date of request for examination] 30.08.2002
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－250938
（P2000－250938A）
(43)公開日　平成12年９月14日(2000.9.14)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'（参考）
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ	15/403	３２０Ｄ　5Ｂ０７５
			15/40	３１０Ｃ
				３４０

審査請求　未請求　請求項の数５　ＯＬ　（全９頁）

(21)出願番号　特願平11－54960

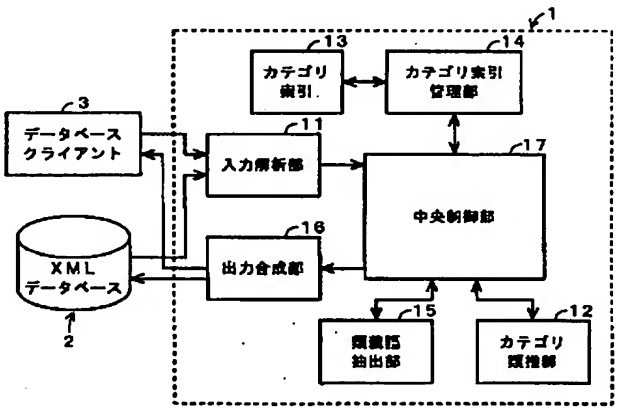
(22)出願日　平成11年３月３日(1999.3.3)

(71)出願人　000001214
ケイディディ株式会社
東京都新宿区西新宿２丁目３番２号
(72)発明者　小野　智弘
埼玉県上福岡市大原２－１－15　株式会社ケイディディ研究所内
(72)発明者　西山　智
埼玉県上福岡市大原２－１－15　株式会社ケイディディ研究所内
(74)代理人　100084870
弁理士　田中　香樹　（外１名）

最終頁に続く

(54)【発明の名称】　ＸＭＬ文書検索装置

(57)【要約】
【課題】　構造上は異なっているが、意味的に類似したＤＴＤをもつＸＭＬデータベースに対して、ユーザがＤＴＤの差異を意識せずに効率的に検索することのできるＸＭＬ文書検索装置を提供することにある。
【解決手段】　データベースクライアント３によって作成された入力用検索式は、入力解析部１１でその要素名が抽出される。中央制御部１７は、該抽出された要素名の類義語を類義語抽出部１５から取得し、該類義語とカテゴリ類推部１２に格納されている要素名とを比較し、一致する要素名を選択する。次に、中央制御部１７は、該選択した要素名を用いて出力用検索式を作り、該出力用検索式を用いてＸＭＬデータベース２を検索する。検索結果は、入力解析部１１、中央制御部１７、および出力合成部１６を介して、データベースクライアント３へ通知される。



【特許請求の範囲】
【請求項１】　複数のＸＭＬ文書から所望の文書を検索するためのＸＭＬ文書検索装置において、
入力された検索式からタグの要素名を抽出する手段と、該抽出された要素名の類義語を抽出する手段と、
該類義語を、ＸＭＬデータベースのタグ定義（ＤＴＤ）に対応したカテゴリ索引と対照し、該カテゴリ索引から前記類義語と一致するタグの要素名を取得する手段と、該カテゴリ索引から取得したタグの要素名を用いて出力用の検索式を作成する手段とを具備し、
該出力用の検索式を用いて、前記ＸＭＬデータベースを検索するようにしたことを特徴とするＸＭＬ文書検索装置。
【請求項２】　請求項１に記載のＸＭＬ文書検索装置において、
前記入力された検索式はルート要素名を有し、該ルート要素名の類義語と一致するタグの要素名を、前記カテゴリ索引から取得するようにしたことを特徴とするＸＭＬ文書検索装置。
【請求項３】　請求項１または２に記載のＸＭＬ文書検索装置において、
前記カテゴリ索引は、カテゴリ名と、それの下位に位置するタグの要素名とからなり、前記入力された検索式のルート要素名の類義語と前記カテゴリ索引のカテゴリ名とが対照され、両者が一致したものについて、さらに該ルート要素名の下位にある要素名の類義語と、前記カテゴリ索引の前記カテゴリ名と関連するタグの要素名との対照がなされるようにしたことを特徴とするＸＭＬ文書検索装置。
【請求項４】　請求項３に記載のＸＭＬ文書検索装置において、
前記カテゴリ索引のカテゴリ名は、前記ＸＭＬデータベースに格納されているタグのデータを基に、類推して決定されることを特徴とするＸＭＬ文書検索装置。
【請求項５】　請求項１～４のいずれかに記載のＸＭＬ文書検索装置において、
前記カテゴリ索引は、前記ＸＭＬデータベースの内容の変化に伴って更新されるようにしたことを特徴とするＸＭＬ文書検索装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】この発明はＸＭＬ文書検索装置に関し、特に、ユーザが検索対象となる文書の型のタグ定義（ＤＴＤ：Document Type Definition）を知らなくても、ＸＭＬ（eXtensible Markup Language）データベースから所望のデータを検索することのできるＸＭＬ文書検索装置に関する。
【０００２】
【従来の技術】近年、インターネットやイントラネット上で文書を記述、交換するための言語として、ＸＭＬが注目されている。ＸＭＬはＨＴＭＬと異なり、構造をもった文書を記述するためのタグを用いることにより、文書を一まとまりではなく、細かい要素の単位で記述、管理することを可能としている。今日までに、ＸＭＬで記述された文書を格納し、検索するためのデータベースが幾つか発表されている。例えば、Object Design 社のeXcelon という名の商品等がある。
【０００３】さて、ＸＭＬ文書では、タグはユーザが自由に定義して使用できるため、全ての利用者間で共通のＤＴＤが利用されるのではなく、情報発信者が独自に定義／拡張したＤＴＤを用いて文書が記述されることがあると考えられる。この結果、インターネットやイントラネット上では、構造上は異なっているが、意味的に類似したＤＴＤをもつＸＭＬ文書が散在することになる。
【０００４】図６に、構造上は異なっているが、意味的に類似した２種類のＤＴＤの例と、これに基づいたＸＭＬ文書および検索式の例を示す。図６(a) は、paper,title,author, およびdateの各タグ（タグの名前を要素名と呼ぶ）を定義しているＤＴＤで、paper が残りの３つを含むことを示している。一方、同図(a')は、article, Title,page, およびwriterの各タグを定義しているＤＴＤで、article が残りの３つを含むことを示している。
【０００５】同図(b) はＸＭＬ文書の表現を示し、paper を起点（ルート要素名）とするＤＴＤに従っていること、各要素名に対する値が、SAMPLE TITLE、John,1103であることを示している。同図(b')は、article を起点とするＤＴＤに従っていること、各要素名に対する値が、SAMPLE TITLE,123,john であることを示している。
【０００６】さらに、同図(c) はＸＭＬ－ＱＬで記述した検索式で、paper をルート要素名とし、authorの値がjohnであるＸＭＬ文書からtitle の値を取得することを示している。また、同図(c')は、article をルート要素名とし、writerの値がjohnであるＸＭＬ文書からTitleの値を取得することを示している。
【０００７】図９は、前記のＸＭＬデータベースを使用した文書の検索例の説明図である。プロセス構成は、ユーザの検索要求を受付け、データベースへデータベース操作言語で要求を送るデータベースクライアント３１と、ＸＭＬ文書を格納し、外部へデータベース操作言語による操作を提供するＸＭＬデータベース３２からなっている。この従来構成では、ユーザあるいはアプリケーションプログラムが、データベースから文書全体あるいはその一部を取得しようとすると、該ユーザ等は目的とする文書が存在しそうな全ての型（例えば、paper 型、article 型）のＤＴＤをそれぞれ理解し、図示されている３３、３４のように、それらの型毎に検索操作を発行することが必要になる。
【０００８】
【発明が解決しようとする課題】前記したように、インターネットやイントラネット上では、構造上は異なって

いるが、意味的に類似したＤＴＤをもつＸＭＬデータベースが散在するため、ユーザあるいはアプリケーションプログラムが、該ＸＭＬデータベースからＸＭＬ文書を検索しようとすると、必要な値があると思われる全てのＤＴＤの文書に対して別々に検索式を記述することが必要になり、効率的でないという問題があった。

【０００９】例えば、図９を例にとると、john氏が書いた著書の題名を知りたい場合、ＸＭＬデータベースでは、paper とarticle で定義される文書は異なったものであるため、paper とarticlのそれぞれに対して、図９の３３、３４のように、別々に検索式を記述して問い合わせることが必要になる。また、このため、そのコストは類似した異なるＤＴＤに基づいて記述された文書が増えるに従って増大するという問題もあった。

【００１０】本発明の目的は、前記した従来技術の問題点を除去し、構造上は異なっているが、意味的に類似したＤＴＤをもつＸＭＬデータベースに対して、ユーザがＤＴＤの差異を意識せずに効率的に検索することのできるＸＭＬ文書検索装置を提供することにある。

【００１１】

【課題を解決するための手段】前記した目的を達成するために、この発明は、ＸＭＬ文書から所望の文書を検索するためのＸＭＬ文書検索装置において、入力された検索式からタグの要素名を抽出する手段と、該抽出された要素名の類義語を抽出する手段と、該類義語を、ＸＭＬデータベースのタグ定義（ＤＴＤ）に対応したカテゴリ索引と対照し、該カテゴリ索引から前記類義語と一致するタグの要素名を取得する手段と、該カテゴリ索引から取得したタグの要素名を用いて出力用の検索式を作成する手段とを具備し、該出力用の検索式を用いて、前記ＸＭＬデータベースを検索するようにした点に特徴がある。

【００１２】この発明によれば、入力された検索式は、該検索式に記述されているタグの要素名の類義語を基に、ＸＭＬデータベース内に実在する文書のタグ定義に対応した要素名をもつ出力用の検索式に自動的に変換されるので、データベースクライアントは検索対象となる文書の型のＤＴＤを知る必要がなく、検索手続きが簡単になると共に、検索範囲を拡張させることができるようになる。

【００１３】

【発明の実施の形態】以下に、図面を参照して、本発明を詳細に説明する。図１は、本発明のＸＭＬ文書検索システムの一実施形態の構成を示すブロック図である。図１に示されているように、ＸＭＬ文書検索システムは、ＸＭＬ文書検索装置１と、ＸＭＬデータベース２と、データベースクライアント３から構成されている。

【００１４】ＸＭＬ文書検索装置１は、外部からの入力を受付けてこれを解析する入力解析部１１と、要素の集合を受取り、その要素の集合を特徴付けるカテゴリ名を出力するカテゴリ類推部１２と、ＸＭＬデータベース２のＤＴＤの情報に対応したカテゴリ索引１３を管理するカテゴリ索引管理部１４と、与えたキーワードの複数の類義語を出力する類義語抽出部１５と、検索装置１の処理結果を外部へ送出する出力合成部１６と、前記各部の全体の制御を行う中央制御部１７から構成されている。

【００１５】前記ＸＭＬ文書検索装置１の構成をさらに詳細に説明すると、前記入力解析部１１は、データベースクライアント３からのデータベース操作要求を受付け、操作要求のパラメタの抽出を行う。また、ＸＭＬデータベース２からの応答を受け付ける。前記カテゴリ類推部１２は同一要素名に属する要素の集合を中央制御部１７から受取り、その要素集合を特徴付けるカテゴリ名を類推し、その中で最も信頼度の高いものを中央制御部１７へ送出する。前記カテゴリ索引管理部１４は、ＸＭＬデータベース２のＤＴＤの情報に対応したカテゴリ索引１３を管理する。

【００１６】前記カテゴリ索引１３は、ＤＴＤのあるタグに対応した要素の集合を特徴付ける「カテゴリ名」を索引鍵とし、それに対応する実際のＤＴＤを値とするものである。該「カテゴリ名」は、実際のＸＭＬデータベース２の値からシソーラスを利用した類推により導出される。

【００１７】また、前記類義語抽出部１５は、与えたキーワードの複数の類義語を出力する。既存のシソーラスＤＢ等が使用可能である。例えば、QZS Dictionary Server等のシソーラスＤＢが使用可能である。前記出力合成部１６は、データベースクライアント３によってなされたデータベース操作要求に伴ってＸＭＬ文書検索装置１によってなされた処理結果である検索式の各パラメタを受取り、複数の検索式を合成してＸＭＬデータベース２に送出する。また、入力解析部１１から転送されたＸＭＬデータベース２からの応答をデータベースクライアント３へ送出する。前記中央制御部１７は、入力解析部１１からパラメタを受取り、カテゴリ類推部１２、カテゴリ索引管理部１４、および類義語抽出部１５を利用して、データベース操作処理、カテゴリ索引構築／変更処理を行い、その結果を出力合成部１６に送る。

【００１８】次に、前記の構成を有するＸＭＬ文書検索装置１の動作を、以下に説明する。まず、該ＸＭＬ文書検索装置１を初めてＸＭＬデータベース２に接続した時に、前記中央制御部１７が行う動作を、図２のフローチャートと図３の具体例を参照して説明する。この動作は、実際のＸＭＬデータベース２の値からカテゴリ索引１３を構築する動作である。

【００１９】ステップＳ１では、ＸＭＬデータベース２から全てのルート要素名と、それに対応する型（ＤＴＤ）を取得し、カテゴリ索引管理部１４へＤＴＤ登録要求を出す。カテゴリ索引管理部１４はカテゴリ索引１３にＤＴＤを登録する。図３の例では、ＸＭＬデータベー

ス2中に格納されているルート要素名「paper 」とそれに対応するDTD「paper,title,author,date 」、次のルート要素名「article」とそれに対応するDTD「article,Title,page,writer 」、さらに次のルート要素名「trip」とそれに対応するDTD「destination,departure,arrival 」、…を、XMLデータベース2から取得し、一旦カテゴリ索引13に登録する。

【0020】ステップS2では、前記ルート要素名の中の、あるルート要素名について、XMLデータベース2から、任意個の文書(data)を取得する。図3の例では、ルート要素名「paper 」に対応する文書「SAMPLE,john,9701」、「SAMPLE2,john,9811 」等を、XMLデータベース2から取得する。

【0021】ステップS3では、取得した複数の文書をカテゴリ類推部12へ送り、送った複数の文書を代表するカテゴリ名を取得する。カテゴリ類推部12では、複数の文書を基にそれを代表するカテゴリ名を類推し、最も信頼度の高いもの(cname)を中央制御部17へ送出する。図3の例では、カテゴリ類推部12が前記文書「SAMPLE,john,9701」、「SAMPLE2,john,9811 」から、カテゴリ名「本」を類推したとする。

【0022】ステップS4では、カテゴリ索引管理部14に対して、該cname の登録要求を出す。カテゴリ索引管理部14は該cname を前記ルート要素名と対応付けてカテゴリ索引13に登録し管理する。図3の例では、cname である「本」をルート要素名「paper 」と関連付けてカテゴリ索引13に登録する。

【0023】ステップS5では、全部のルート要素名にcname が対応付けられたか否かの判断がなされ、この判断が否定の時にはステップS2に戻って、前記の動作が繰り返される。図3の例では、次に、ルート要素名「article 」に対応する文書「Flower,101,thomas 」、「Animals,100,tom 」、「Database,56,john」が取得され、これらから例えばカテゴリ名「本」が類推されて、cname である「本」をルート要素名「article 」と関連付けてカテゴリ索引13に登録する。

【0024】以上の処理が繰返し行われ、前記ステップS5の判断が肯定になると、カテゴリ索引構築の処理は終了する。以上の動作により、例えば、図5に示されているような、カテゴリ索引13が作成される。

【0025】なお、構築されたカテゴリ索引は、データ型の挿入や更新に伴って変更したり、格納する文書の増加あるいは変化に伴ってカテゴリ名の精度を向上させる等により、維持することが必要である。このカテゴリ名の更新は、データ操作やデータ型操作を契機として、前記中央制御部17とカテゴリ索引管理部14とカテゴリ類推部12が行う。

【0026】次に、XML文書検索装置1のデータ検索処理の動作を、図4のフローチャートおよび図5の説明図を参照して説明する。ステップS11では、前記データベースクライアント3の検索操作により、検索式の入力があったか否かの判断がなされる。この判断が肯定になるとステップS12に進み、ある数iが1と置かれる。ステップS13では、前記検索式21から、ルート要素名と、パラメタ要素名と、その値が抽出される。抽出されたパラメタ数(ルート要素名+パラメタ要素名)の個数をx個とする。

【0027】例えば、図5に示されているように、データベースクライアント3から、検索式21が入力されたとすると、該検索式は入力解析部11を通って中央制御部17に送られる。該中央制御部17は、検索式21から、ルート要素名「文書」と、パラメタの要素名に相当する「著者」とその値である「john」と、他の要素名である「題名」を抽出する。この場合には、パラメタ数x=3となる。

【0028】ステップS14では、類義語抽出部15へ、該抽出したルート要素名とパラメタの要素名を渡し、それぞれの類義語を取得する。図5の例では、ルート要素名である「文書」と、パラメタの要素名である「著者」と「題名」が、類義語抽出部15に渡される。そうすると、該類義語抽出部15は、前記ルート要素名およびパラメタの要素名に対応する類義語を中央制御部17に回答する。なお、該類義語抽出部15としては、市販のシソーラスDB23を使用することができる。

【0029】ステップS15では、該ルート要素名の類義語、例えば前記「文書」の類義語である本、paper,Paper,Document,article等を前記カテゴリ索引管理部14を通してカテゴリ索引13に送り、該カテゴリ索引13から、該類義語をカテゴリ名にもつルート要素名とDTDを取得する。図5の例では、カテゴリ索引13から、カテゴリ索引「本」に対応するルート要素名「paper 」と「article 」とを取得する。また、各ルート要素名に対応するDTDを取得する。

【0030】ステップS16では、カテゴリ索引の中に、前記ルート要素名の類義語群が存在するか否かの判断がなされる。この判断が否定の時には、処理を終了する。一方、肯定の時には、ステップS17に進んで、前記カテゴリ索引から取得したルート要素名の個数をk個とし、i番目のルート要素名のDTDを取得し、該DTDの中で前記類義語と一致する要素名を選択する。この時、選択した要素名の個数をyとする。

【0031】図5の例では、ルート要素名「paper 」のDTD「paper,title,author,date」を取得し、前記ルート要素名の下位のパラメタの類義語「author,writer,Author,....,Title,title,Theme,... 」と一致する要素名を、前記DTDから選択する。この例では、「paper,title,author」が一致するので、該「paper,title,author」が選択される。

【0032】ステップS18では、該一致した要素名の個数y=前記検索式から抽出したパラメタ個数xが成立

するか否かの判断を行い、この判断が肯定の場合には、ステップＳ１９に進んで、出力検索式を１個作成する。図５の例では、「paper,title,author」を用いて一つの出力検索式が作成される。

【００３３】ステップＳ２０では、ｉ≧ｋが成立するか否かの判断が行われる。この判断が否定の時およびステップＳ１８の判断が否定の時には、ステップＳ２１に進んでｉに１が加算される。そして、ステップＳ１７に戻って、次のルート要素名（図５の例では、「article」）のＤＴＤを取得し、該ＤＴＤの中で前記類義語と一致する要素名を選択する。この例では、「article,writer,Title」が選択される。以上の動作が繰返し行われ、ステップＳ２０の判断が肯定になると、ステップＳ２２に進んで、前記出力合成部１６にて、出力検索式の合成が行われる。図５の例では、この合成により、出力検索式２２ａと２２ｂが得られることになる。

【００３４】ステップＳ２３では、該検索式２２ａと２２ｂが前記ＸＭＬデータベース２に送られる。ステップＳ２４では、ＸＭＬデータベース２からの応答が収集されて入力解析部１１を介して出力合成部１６に送られ、ステップＳ２５では収集結果が該出力合成部１６からデータベースクライアント３へ送られる。

【００３５】以上のようにして、上記の実施形態によれば、ユーザはＤＴＤの要素名の差や配置を意識せずに、ＸＭＬデータベースを効率的に検索することができるようになる。

【００３６】次に、本発明の第２実施形態を、図６および図７を参照して説明する。図６は前記カテゴリ索引１３を構築する動作の説明図である。この実施形態は、図３で示したようなカテゴリ類推部１２を用いずに、ＸＭＬデータベース２から、この中に格納されているルート要素名とそれに対応するＤＴＤを任意の個数または全部取得し、カテゴリ索引１３に登録するようにしたものである。この方法によれば、図７に示されているような内容の、ルート要素名とＤＴＤがカテゴリ索引１３として登録されることになる。

【００３７】次に、ＸＭＬ文書検索装置１のデータ検索処理の動作を図７を参照して説明する。本実施形態の動作が図５の動作と異なる点は、中央制御部１７が、類義語抽出部１５から取得したルート要素名の類義語を基に、カテゴリ索引１３のルート要素名を検索するようにしたことにあり、他の点は、図５と同じである。

【００３８】この実施形態によれば、ＸＭＬデータベースの検索の精度は、前記第１実施形態に比べて若干低下すると考えられるが、カテゴリ索引１３を簡単な構成でかつ安価に構築できるという利点を有している。

【００３９】

【発明の効果】以上の説明から明らかなように、本発明によれば、入力された検索式からタグの要素名を抽出し、該要素名を、その類義語を基にＸＭＬデータベースに格納されているタグの要素名に変換して、出力検索式を作成するようにしているので、ユーザは、検索対象となるＸＭＬデータベースの文書の型のＤＴＤを予め知っている必要がなく、簡単に検索式を作成することができる。したがって、ユーザは効率的に検索でき、しかも、精度良く検索結果を取得することができる。

【００４０】また、カテゴリ索引は、ＸＭＬデータベースの文書に追加、変更、削除等の更新があると自動的に更新されるので、何らのメンテナンスをすることなく、最良の状態に維持できる。

【図面の簡単な説明】

【図１】　本発明の一実施形態の概略の構成を示すブロック図である。

【図２】　本発明の第１実施形態のカテゴリ索引構築の動作を示すフローチャートである。

【図３】　該第１実施形態のカテゴリ索引構築の動作説明図である。

【図４】　本発明の第１実施形態のＸＭＬ文書検索装置のデータ検索処理の動作を示すフローチャートである。

【図５】　前記第１実施形態のＸＭＬ文書検索装置のデータ検索処理の動作説明図である。

【図６】　本発明の第２実施形態のカテゴリ索引構築の動作説明図である。

【図７】　本発明の第２実施形態のＸＭＬ文書検索装置のデータ検索処理の動作説明図である。

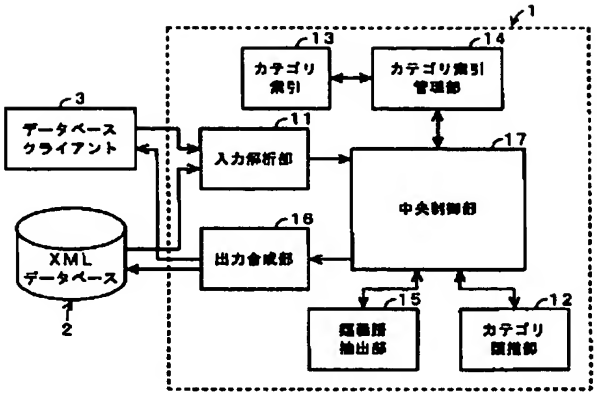
【図８】　ＤＴＤ、ＸＭＬ文書、および検索式の一例の説明図である。

【図９】　従来のＸＭＬ文書検索方法の説明図である。

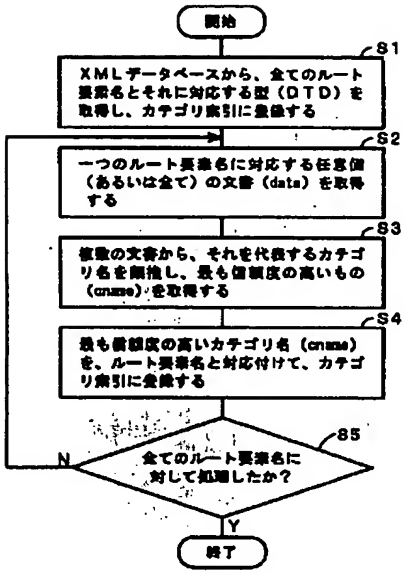
【符号の説明】

１…ＸＭＬ文書検索装置、２…ＸＭＬデータベース、３…データベースクライアント、１１…入力解析部、１２…カテゴリ類推部、１３…カテゴリ索引、１４…カテゴリ索引管理部、１５…類義語抽出部、１６…出力合成部、２１…入力された検索式、２２ａ，２２ｂ…出力検索式。

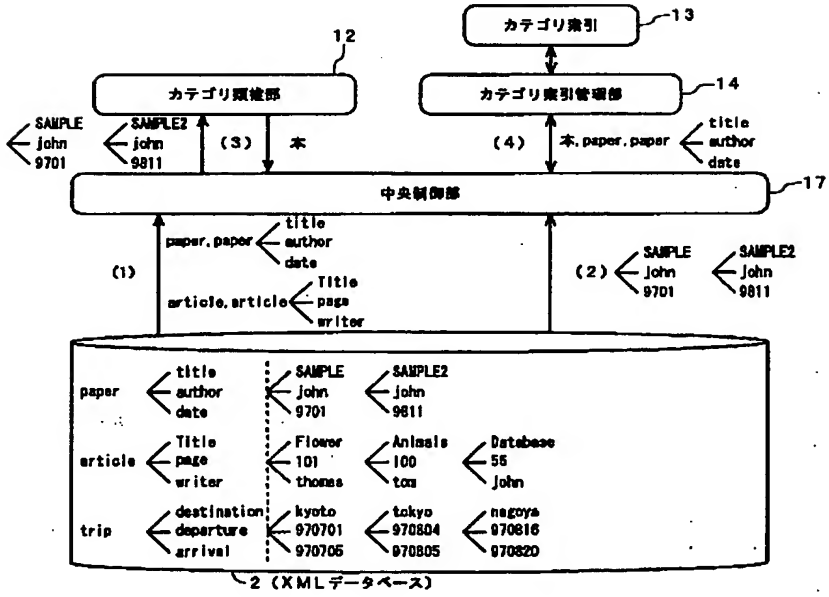
【図1】


1
13
カテゴリ索引
14
カテゴリ索引管理部
3
データベースクライアント
11
入力解析部
17
中央制御部
16
出力合成部
XML データベース
2
15
属性語抽出部
12
カテゴリ推定部

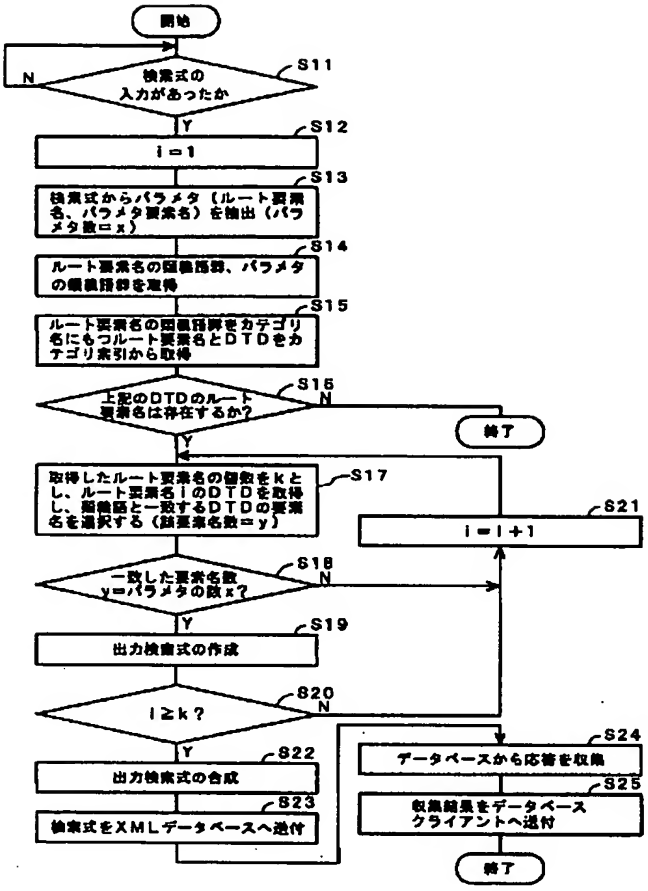
【図2】


開始
S1
XMLデータベースから、全てのルート要素名とそれに対応する型（DTD）を取得し、カテゴリ索引に登録する
S2
一つのルート要素名に対応する任意個（あるいは全て）の文書（data）を取得する
S3
複数の文書から、それを代表するカテゴリ名を類推し、最も信頼度の高いもの（cname）を取得する
S4
最も信頼度の高いカテゴリ名（cname）を、ルート要素名と対応付けて、カテゴリ索引に登録する
S5
全てのルート要素名に対して処理したか？
N
Y
終了

【図3】


12
カテゴリ類推部
13
カテゴリ索引
14
カテゴリ索引管理部
SAMPLE
john
9701
SAMPLE2
john
9811
(3)
本
(4)
本, paper, paper
title
author
date
17
中央制御部
(1)
paper, paper
title
author
date
article, article
Title
page
writer
(2)
SAMPLE
John
9701
SAMPLE2
John
9811
paper
title
author
date
SAMPLE
john
9701
SAMPLE2
john
9811
article
Title
page
writer
Flower
101
thomas
Animals
100
tom
Database
55
John
trip
destination
departure
arrival
kyoto
970701
970705
tokyo
970804
970805
nagoya
970816
970820
2 (XMLデータベース)

【図4】



【図8】

(a) DTDの例1

```
<!ELEMENT paper (title, author, date)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT date (#PCDATA)>
```

(b) XML文書の例1

```
<?xml version="1.0"?>
<!DOCTYPE paper
  SYSTEM "http://www.a.b.c/pap.dtd">
<paper><title> SAMPLE TITLE </>
<author> john </><date>1103</></>
```

(c) XML-QLで記述した検索式の例1

```
WHERE <paper>
        <author> john </>
        <title> $a </>
      </> IN "www.a.b.c/bib.xml"
CONSTRUCT $a
```

(a') DTDの例2

```
<!ELEMENT article (Title, page, writer)>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT page (#PCDATA)>
<!ELEMENT writer (#PCDATA)>
```

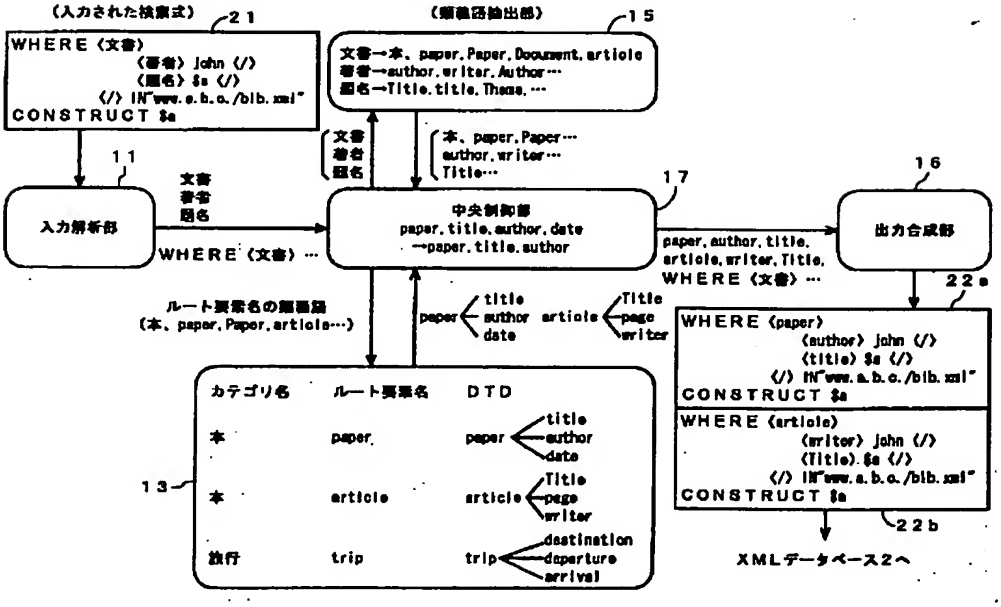
(b') XML文書の例2

```
<?xml version="1.0"?>
<!DOCTYPE article
  SYSTEM "http://www.a.b.c/pap.dtd">
<article><Title> SAMPLE TITLE </>
<page> 123 </><writer> john</></>
```

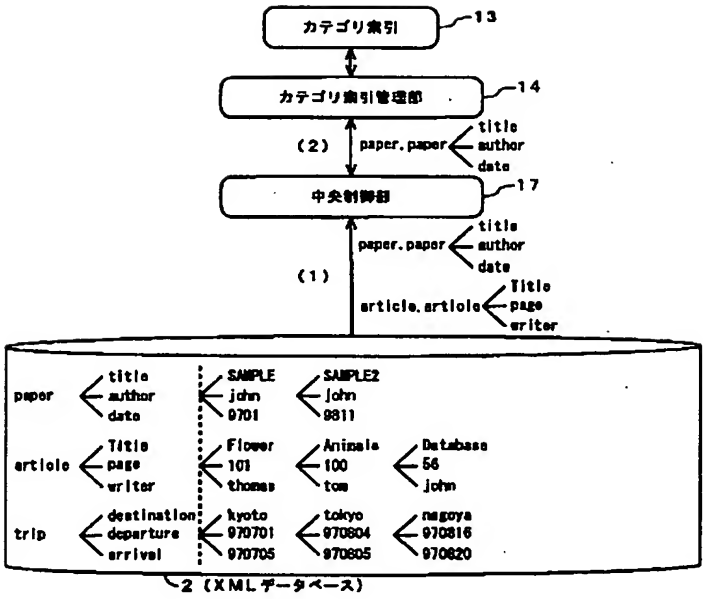
(c') XML-QLで記述した検索式の例2

```
WHERE <article>
        <writer> john </>
        <Title> $a </>
      </> IN "www.a.b.c/bib.xml"
CONSTRUCT $a
```

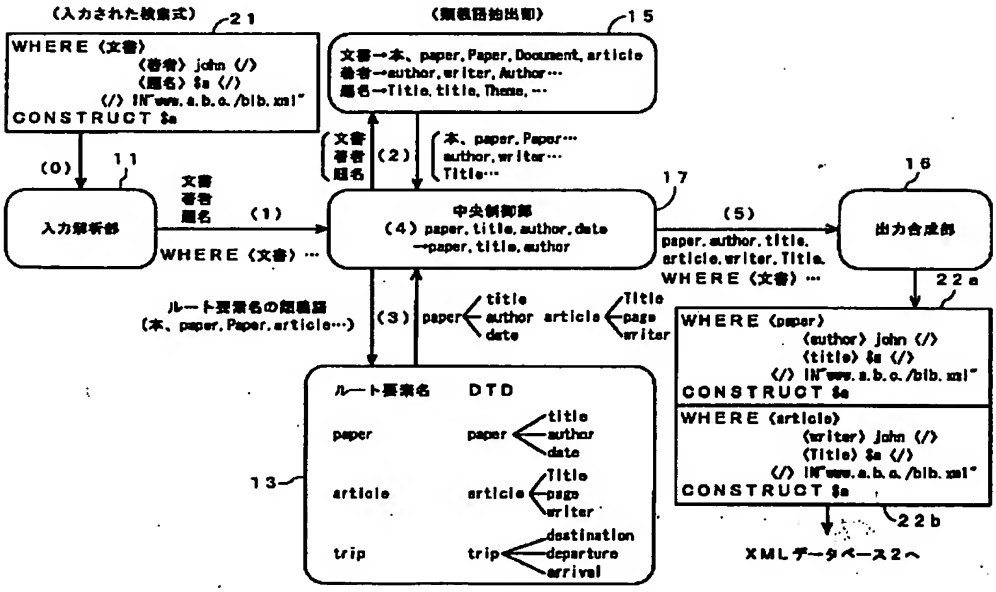
【図5】



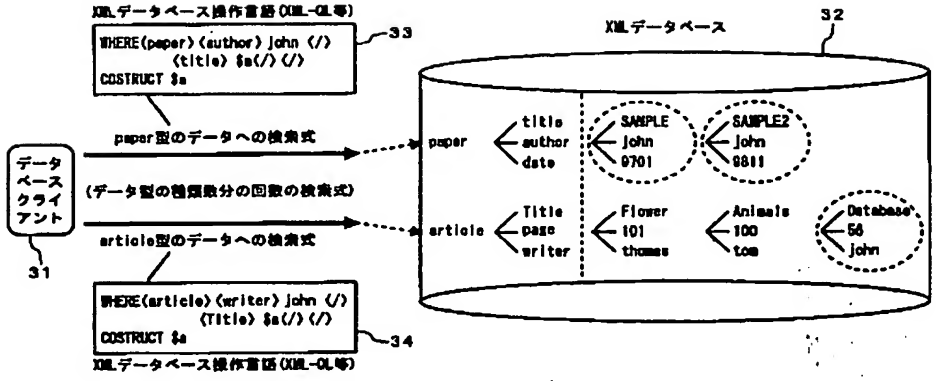
【図6】


カテゴリ索引
13
カテゴリ索引管理部
14
(2)
paper, paper
title
author
date
中央制御部
17
(1)
paper, paper
title
author
date
article, article
Title
page
writer
paper
title
author
date
SAMPLE
john
9701
SAMPLE2
john
9811
article
Title
page
writer
Flower
101
thomas
Animals
100
tom
Database
56
john
trip
destination
departure
arrival
kyoto
970701
970705
tokyo
970804
970805
nagoya
970816
970820
2 (XMLデータベース)

【図7】


(入力された検索式)
21
WHERE (文書)
(著者) john (/)
(題名) $a (/)
(/) IN"www.a.b.c./bib.xml"
CONSTRUCT $a
(類義語抽出部)
15
文書→本、paper, Paper, Document, article
著者→author, writer, Author…
題名→Title, title, Theme, …
(0)
11
入力解析部
文書
著者
題名
(1)
WHERE (文書) …
(2)
文書
著者
題名
本、paper, Paper…
author, writer…
Title…
中央制御部
17
(4) paper, title, author, date
→paper, title, author
(5)
paper, author, title,
article, writer, Title,
WHERE (文書) …
16
出力合成部
22a
WHERE (paper)
(author) john (/)
(title) $a (/)
(/) IN"www.a.b.c./bib.xml"
CONSTRUCT $a
WHERE (article)
(writer) john (/)
(Title) $a (/)
(/) IN"www.a.b.c./bib.xml"
CONSTRUCT $a
22b
XMLデータベース2へ
ルート要素名の候補語
(本、paper, Paper, article…)
(3)
paper
title
author
date
article
Title
page
writer
13
ルート要素名
DTD
paper
paper
title
author
date
article
article
Title
page
writer
trip
trip
destination
departure
arrival

【図9】



フロントページの続き

(72)発明者　小花　貞夫
埼玉県上福岡市大原２－１－15　株式会社
ケイディディ研究所内

Fターム(参考)　5B075 KK02 KK07 KK13 KK37 KK39
ND03 ND35 NK02 NK32 NK35
PP23 PP25 PP26 PR06 QM07
QP03 UU06